THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 56, No. 3, March 1977
Printed in U.S.A.

Rate vs Fidelity for the Binary Source

By S. P. LLOYD

(Manuscript received May 5, 1976)

Errors are deliberately introduced in the output of a binary message source to reduce the entropy rate. The errors depend on the source sequence in a deterministic shift-invariant manner. The tradeoff between error rate permitted and reduction of entropy rate is of interest. It is shown that the ideal bound cannot be attained. If the errors are required to be produced causally, then a bound stronger than the ideal bound takes over. The quantities of interest are found explicitly for the example: change all 0's in 0-runs of length 1 to 1's.

If a transmission channel has capacity C bits/second and a message source has entropy rate H bits/second satisfying $H \leqq C$, then the source can be encoded, fed to the channel, decoded at the channel output, and recovered essentially without error after such handling. The rate-distortion theory is concerned with the case where $H > C$; we try to minimize some measure of the errors that are necessarily present.[1]

We treat here a special class of systems in which the errors are deliberately introduced before submission to the channel to reduce the entropy rate to that of the channel; the mutilated source is then handled without further error by the channel. The usual treatment involves use of block codes, but we will examine the more interesting sliding (or shift-invariant) codes.

The source in Fig. 1 emits letters x_n, $-\infty < n < \infty$, at rate 1 per unit time. The letters are drawn from alphabet $A = \{0,1\}$ with probability distribution $P\{x_n = 0\} = P\{x_n = 1\} = 1/2$, the same for all n, and the draws are statistically independent. We denote by $x = (x_n: -\infty < n < \infty)$ a sample sequence of the source process X. The entropy rate of the source is $H(X) = 1$ bit per unit time.

The error generator operates on a source sequence x to produce a sequence $e = (e_n: -\infty < n < \infty)$ of A valued random variables $e_n = e_n(x)$. The error at time n is a deterministic function $e_n = \eta(\cdots, x_{n-1}; x_n; x_{n+1}, \cdots)$ of the whole sample sequence x. The measurable function η is the same for all n, so that the dependence of e on x is shift

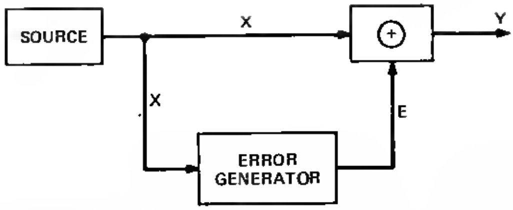

Fig. 1—Reducing the entropy rate by introducing errors.

invariant: if sequence x is shifted m places, the sequence $e = e(x)$ shifts m places with it.

The output of the adder box "$\oplus$" in Fig. 1 is simply $y_n = x_n \oplus e_n$, with $\oplus$ the usual addition mod 2. We regard the output process Y as X corrupted by the errors E. Now, depending on how E is generated, process Y can have entropy rate $H(Y) < H(X)$, and so can be handled by a channel of correspondingly smaller capacity at the price of the errors introduced. We are concerned with the tradeoff between the error rate and the decrease in entropy rate. Explicitly, suppose error rate ϵ is specified, $0 \leqq \epsilon \leqq 1/2$, and that $\eta(\cdots ;; \cdots)$ is a stationary error-generating function with the property $P\{e_n = 1\} = \epsilon$. The resulting Y process will have a certain entropy rate $H(Y) \leqq H(X)$ determined by η. What is the least value that $H(Y)$ can have for all such η?

I. THE IDEAL BOUND

Let us consider the joint process $Z = (Y,E)$, where the Z alphabet is $\{(0,0),(0,1),(1,0),(1,1)\}$ and each z_n in a sequence $z = (z_n: -\infty < n < \infty)$ is the pair $z_n = (y_n,e_n)$. The mapping $\Theta: X \to Z$, which sends a sample sequence x to sequence $z = \Theta x$, is obviously shift invariant. The map Θ is also measure preserving by definition; the probability measure on the space of sequences z is that induced by Θ and the X distribution. In the other direction, $x_n = y_n \oplus e_n$, $-\infty < n < \infty$ is the inverse map $\Phi: Z \to X$, which recovers the source sequence x if the compressed version y and tbe errors e are known. Tbis map is also shift invariant and measure preserving. Since processes X and Z are isomorphic in the above sense, their entropy rates are the same: $H(Z) = H(X) = 1$.

From the general theory (Section 6 in Ref. 3), the entropy rate $H(Z)$ is the average conditional entropy

$$
\begin{aligned}
H(Z) &= H(z_1|\cdots,z_0) \\
&= H[(y_1,e_1)|\cdots,(y_0,e_0)]
\end{aligned}
$$

of letter z_1, given the preceding letters $\cdots,z_0$. Using the addition law for conditional entropy, we find

$$H(Z) = H[e_1|\cdots,(y_0,e_0)] + H[y_1|\cdots,(y_0,e_0) \text{ and } e_1]$$
$$\leqq H(e_1) + H(y_1|\cdots,y_0)$$
$$= h(\epsilon) + H(Y),$$

since $H(e_1) = h(\epsilon) \triangleq \epsilon \log_2(1/\epsilon) + (1-\epsilon)\log_2[1/(1-\epsilon)]$ when $P\{e_1 = 1\} = \epsilon$, $P\{e_1 = 0\} = 1-\epsilon$. Using $H(Z) = 1$, we have the lower bound $H(Y) \geqq 1 - h(\epsilon)$, $0 \leqq \epsilon \leqq 1/2$, for any such compression scheme.

Our first result is:

Theorem 1: *For error rate* $0 < \epsilon < 1/2$ *it is not possible to attain the bound* $H(Y) = 1 - h(\epsilon)$.

Proof: For each fixed $N \geqq 1$, there holds $NH(Z) = H(z_1,\cdots,z_N|\cdots,z_0)$, by induction from

$$H(z_1,\cdots,z_N|\cdots,z_0) = H(z_1|\cdots,z_0) + H(z_2,\cdots,z_N|\cdots,z_1).$$

Arguing as before, we find

$$N = H[(y_1,e_1),\cdots,(y_N,e_N)|\cdots,(y_0,e_0)]$$
$$= H[y_1,\cdots,y_N|\cdots,(y_0,e_0)]$$
$$+ H[e_1,\cdots,e_N|\cdots,(y_0,e_0) \text{ and } y_1,\cdots,y_N];$$

moreover,

$$H[y_1,\cdots,y_N|\cdots,(y_0,e_0)] \leqq H(y_1,\cdots,y_N|\cdots,y_0) = NH(Y)$$

and

$$H(e_1,\cdots,e_N|\cdots,e_0 \text{ and } \cdots,y_N) \leqq NH(e_1) = Nh(\epsilon).$$

Now, equality in this last step holds *iff* $e_1,\cdots,e_N, f(\cdots,e_0 \text{ and } \cdots,y_N)$ are mutually independent, f is any measurable function of the variables indicated. (Equality in the first step requires that $y_1,\cdots,y_N$ be conditionally independent of $\cdots,e_0$ given $\cdots,y_0$, but we will not need this.)

For real valued variable u, let us define

$$u^{(\alpha)} = \begin{cases} u & \text{if } \alpha = 0, \\ 1-u & \text{if } \alpha = 1; \end{cases}$$

we put also $u^{(\alpha)(\beta)} = [u^{(\alpha)}]^{(\beta)} = [u^{(\beta)}]^{(\alpha)}$ for all $\alpha,\beta \epsilon A$; note that $u^{(0)(0)} = u^{(1)(1)} = u$, $u^{(0)(1)} = u^{(1)(0)} = 1-u$. From

$$x_j = y_j \oplus e_j$$
$$= y_j e_j + (1-y_j)(1-e_j)$$
$$= y_j^{(0)} e_j^{(0)(0)} + y_j^{(1)} e_j^{(1)(0)}$$

and

$$\begin{aligned}1 - x_j &= y_j \oplus (e_j \oplus 1)\\ &= y_j(1 - e_j) + (1 - y_j)e_j\\ &= y_j^{(0)}e_j^{(0)(1)} + y_j^{(1)}e_j^{(1)(1)},\end{aligned}$$

it is apparent that

$$x_j^{(\alpha)} = \sum_\beta y_j^{(\beta)}e_j^{(\beta)(\alpha)},$$

where α,β are variables in the set A. Multiplying these equations together for $1 \leqq j \leqq N$ gives

$$x_1^{(\alpha_1)} \cdots x_N^{(\alpha_N)} = \sum_{\beta_1} \cdots \sum_{\beta_N} y_1^{(\beta_1)} \cdots y_N^{(\beta_N)} e_1^{(\alpha_1)(\beta_1)} \cdots e_N^{(\alpha_N)(\beta_N)},$$

for each of the 2^N choices for $\alpha_1, \cdots, \alpha_N$.

If $H(Y) = 1 - h(\epsilon)$, then $e_1, \cdots, e_N, [y_1^{(\beta_1)} \cdots y_N^{(\beta_N)}]$ are mutually independent for each choice of the β's. Since $E\{e_j^{(\gamma)}\} = \epsilon^{(\gamma)}$, $1 \leqq j \leqq N$, we find

$$\begin{aligned}\frac{1}{2^N} &= E\{x_1^{(\alpha_1)} \cdots x_N^{(\alpha_N)}\}\\ &= \sum_{\beta_1} \cdots \sum_{\beta_N} \epsilon^{(\alpha_1)(\beta_1)} \cdots \epsilon^{(\alpha_N)(\beta_N)} E\{y_1^{(\beta_1)} \cdots y_N^{(\beta_N)}\}, \text{ all } \alpha\text{'s}.\end{aligned}$$

Using now the assumption $\epsilon \neq 1/2$, let c be the number $c = -\epsilon/(1 - 2\epsilon)$, so that $c^{(1)} = (1 - \epsilon)/(1 - 2\epsilon)$. From

$$\sum_\alpha c^{(\alpha)(\gamma)} = 1, \quad \sum_\alpha c^{(\alpha)(\gamma)}\epsilon^{(\alpha)(\beta)} = \delta_{\gamma,\beta},$$

we obtain

$$\begin{aligned}\frac{1}{2^N} &= \sum_{\alpha_1} \cdots \sum_{\alpha_N} c^{(\alpha_1)(\gamma_1)} \cdots c^{(\alpha_N)(\gamma_N)} \times \frac{1}{2^N}\\ &= \sum_{\alpha_1} \cdots \sum_{\alpha_N} \sum_{\beta_1} \cdots \sum_{\beta_N} c^{(\alpha_1)(\gamma_1)}\epsilon^{(\alpha_1)(\beta_1)} \cdots c^{(\alpha_N)(\gamma_N)}\epsilon^{(\alpha_N)(\beta_N)}\\ &\qquad\qquad \times E\{y_1^{(\beta_1)} \cdots y_N^{(\beta_N)}\}\\ &= E\{y_1^{(\gamma_1)} \cdots y_N^{(\gamma_N)}\}, \quad \text{all } \gamma\text{'s}.\end{aligned}$$

If this holds for all $N \geqq 1$, then the $\{y_n: -\infty < n < \infty\}$ are independent identically distributed random variables with distribution $P\{y_n = 0\} = P\{y_n = 1\} = 1/2$. The entropy rate of this process is $H(Y) = 1 \neq 1 - h(\epsilon)$. □

II. THE CAUSAL BOUND

We now consider the case where each e_n depends only on the present and past values of the x's. That is, $e_n = \eta(\cdots, x_{n-1}; x_n)$, $-\infty < n < \infty$, for η a measurable function of the variables indicated. The relation between Z and X is thus bicausal: z_n depends only on $\cdots, x_n$ and x_n depends only on $\cdots, z_n$. It follows that conditionals given $\cdots, z_n$ agree w.p.1 with conditionals given $\cdots, x_n$.

Theorem 2: If the dependence of the error process E on X is causal, then $H(Y) \geqq 1 - 2\epsilon, 0 \leqq \epsilon \leqq 1/2$.

Proof: Setting A variant form of the basic inequality is $H(Y) \geqq 1 - H(e_0|\cdots, x_{-1})$, obtained from

$$1 = H(Z) = H[(y_0, e_0)|\cdots, (y_{-1}, e_{-1})]$$
$$= H[e_0|\cdots (y_{-1}, e_{-1})] + H[y_0|\cdots, (y_{-1}, e_{-1}) \text{ and } e_0]$$
$$\leqq H(e_0|\cdots, x_{-1}) + H(y_0|\cdots, y_{-1});$$

we have used only that $y_n \oplus e_n$ is less informative than (y_n, e_n), $-\infty < n \leqq -1$. The assumption that η is causal is not involved.

Let us partition the space of sample sequences x into the four disjoint subsets:

$$A_1 = \{x: \eta(\cdots, x_{-1}; 0; x_1, \cdots) = 0 \text{ and } \eta(\cdots, x_{-1}; 1; x_1, \cdots) = 0\}$$
$$A_2 = \{x: \eta(\cdots, x_{-1}; 0; x_1, \cdots) = 1 \text{ and } \eta(\cdots, x_{-1}; 1; x_1, \cdots) = 1\}$$
$$A_3 = \{x: \eta(\cdots, x_{-1}; 0; x_1, \cdots) = 0 \text{ and } \eta(\cdots, x_{-1}; 1; x_1, \cdots) = 1\}$$
$$A_4 = \{x: \eta(\cdots, x_{-1}; 0; x_1, \cdots) = 1 \text{ and } \eta(\cdots, x_{-1}; 1; x_1, \cdots) = 0\}.$$

The random variable $\kappa(x)$ is defined as the part namer for this partition; i.e., $\kappa(x) = j$ *iff* $x \in A_j$, $1 \leqq j \leqq 4$. Since $\kappa(x)$ depends only on coordinates $\cdots, x_{-1}$ and $x_1, \cdots$ of x, the conditional distribution of x_0 given κ is $P\{x_0 = 0|\kappa\} = P\{x_0 = 1|\kappa\} = 1/2$ w.p.1. The resulting random conditional entropies of e_0, y_0 are seen to be

$$h(e_0|\cdots, x_{-1} \text{ and } x_1, \cdots) = h(e_0|\kappa)$$
$$= \begin{cases} 0 \text{ for } \kappa = 1,2 \\ 1 \text{ for } \kappa = 3,4 \end{cases}$$
$$h(y_0|\cdots, x_{-1} \text{ and } x_1, \cdots) = h(y_0|\kappa)$$
$$= \begin{cases} 1 \text{ for } \kappa = 1,2 \\ 0 \text{ for } \kappa = 3,4. \end{cases}$$

Putting $a_i = P\{A_i\}, 1 \leqq i \leqq 4$, the average conditional entropies are then

$$H(e_0|\cdots,x_{-1} \text{ and } x_1,\cdots) = H(e_0|\kappa) = a_3 + a_4$$

$$H(y_0|\cdots,x_{-1} \text{ and } x_1,\cdots) = H(y_0|\kappa) = a_1 + a_2.$$

The error rate is

$$\epsilon = P\{e_0 = 1\} = \frac{1}{2}(a_3 + a_4) + a_2,$$

so we have

$$H(e_0|\cdots,x_{-1} \text{ and } x_1,\cdots) = 2\epsilon - 2a_2 \leqq 2\epsilon.$$

Assume now that E depends causally on X; then, e_0 is conditionally independent of $x_1,\cdots$ given $\cdots,x_{-1}$, implying $H(e_0|\cdots,x_{-1} \text{ and } x_1,\cdots) = H(e_0|\cdots,x_{-1})$. Combining the inequalities, we obtain $H(Y) \geqq 1 - 2\epsilon$. □ This bound is strictly above the ideal bound when $0 < \epsilon < 1/2$, since $h(\epsilon) > 2\epsilon$ on this interval.

III. EXAMPLE

The following example is mentioned in Ref. 4, but a solution is not given. Let the errors be $e_n = \eta(x_{n-1};x_n;x_{n+1})$, $-\infty < n < \infty$, with η the function

$$\eta(1;0;1) = 1$$

$$\eta(x_{-1};x_0;x_1) = 0 \text{ if } x_{-1}x_0x_1 \neq 101.$$

The error rate is $P\{e_n = 1\} = 1/8$. We will compute $H(Y)$ and $H(E)$ explicitly and compare $H(Y)$ with the bounds of Sections I and II.

A graphical representation of η is given in Fig. 2. The vertices of the directed graph are the state pairs $x_{-1}x_0$, the arrows represent the transitions from $x_{-1}x_0$ to x_0x_1, and the value $\eta(x_{-1};x_0;x_1)$ is shown on the arrow from $x_{-1}x_0$ to x_0x_1. The corresponding graph of $y_0 = x_0 \oplus \eta(x_{-1};x_0;x_1)$ appears in Fig. 3.

We now compute $H(Y)$. Examination of Fig. 3 reveals that process Y is a renewal process, with renewal at the beginning of each run, either

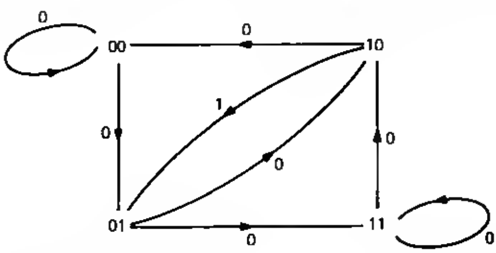

Fig. 2—$e_0 = \eta(x_{-1};x_0;x_1)$. Values of e_0 as function of the transition.

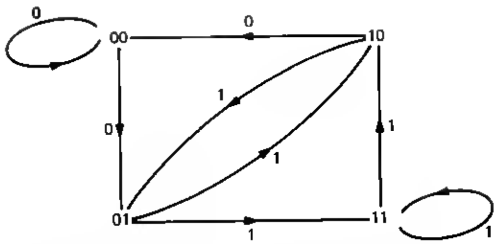

Fig. 3—$y_0 = x_0 \oplus \eta(x_{-1};x_0;x_1)$. Values of y_0 as function of the transition.

of 0's or of 1's. Moreover, the length $R^{(0)}$ of a 0-run has the geometric distribution

$$P\{R^{(0)} = j\} = \frac{1}{2^{j-1}}, \quad j = 2,3,\cdots.$$

The mean and entropy of this distribution are easily found to be $E\{R^{(0)}\} = 3, H(R^{(0)}) = 2$.

The 1-runs of Y involve the subgraph shown in Fig. 4, relabelled for convenience. A 1-run results from a path (driven by x) which starts at A and follows lettered arrows until exit occurs at B along the dotted arrow. If the length $R^{(1)}$ of the run has value $R^{(1)} = j$, the driving x's have probability $2^{-(j+1)}$ per path, so

$$P\{R^{(1)} = j\} = \frac{\nu_j}{2^{j+1}}, \quad j = 1,2,\cdots,$$

where ν_j is the number of paths of length j from A to B along lettered arrows.

For $j \geqq 4$, we classify the paths of length j from A to B according to the earliest appearance of arrow a:

(*i*) One path $c(d)_{j-2}e$ which does not contain a.

(*ii*) Paths which start $ba\cdots$.

(*iii*) For each $0 \leqq k \leqq j-4$, paths which start $c(d)_k ea\cdots$.

In (*ii*) the continuations "$\cdots$" are just the paths from A to B of length

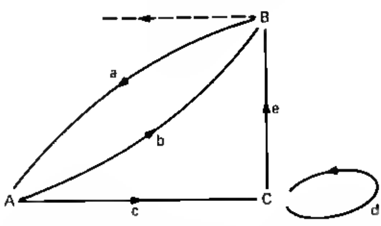

Fig. 4—Subgraph for 1-runs of process Y.

$j-2$, one each, and in (*iii*) the lengths of the continuations are $j-(k+3)$ for each $0 \leqq k \leqq j-4$. In consequence,

$$\nu_j = 1 + \nu_{j-2} + \sum_{k=0}^{j-4} \nu_{j-(k+3)}, \quad j \geqq 4.$$

The initial terms for the recursion are $\nu_1 = 1$, $\nu_2 = 1$,$\nu_3 = 2$, clearly, and it is convenient to define $\nu_0 = 0$. Then from

$$\begin{aligned} \nu_j &= 1 + \nu_{j-2} + \nu_{j-3} + \cdots + \nu_0, \quad & j \geqq 2, \\ \nu_{j-1} &= 1 + \nu_{j-3} + \cdots + \nu_0, & j \geqq 3, \end{aligned}$$

it is apparent that

$$\nu_j = \nu_{j-1} + \nu_{j-2}, \quad j \geqq 3.$$

That is, $\nu_1,\nu_2, \cdots$ is the Fibonacci sequence 1,1,2,3,5,8, $\cdots$.

The generating function for the Fibonacci sequence is $\Sigma_1^\infty \nu_j x^j = x/(1-x-x^2)$, as is well known, so the distribution of $R^{(1)}$ has generating function

$$\sum_{j=1}^{\infty} x^j P\{R^{(1)} = j\} = \frac{x}{4-2x-x^2}, \quad |x| \leqq 1 < \sqrt{5} - 1.$$

Taking $(d/dx)_{x=1}$, we obtain $E\{R^{(1)}\} = 5$ for the mean length of 1-runs in the Y process. For numerical evaluation of the entropy $H(R^{(1)})$ of the $R^{(1)}$ distribution, we obtain the $r_j = P\{R^{(1)} = j\}$, $j \geqq 1$, from the recursion

$$r_j = \frac{1}{2} r_{j-1} + \frac{1}{4} r_{j-2}, \quad j \geqq 3;$$

$$r_1 = \frac{1}{4}, \quad r_2 = \frac{1}{8}.$$

The numerical result is

$$\begin{aligned} H(R^{(1)}) &= \sum_{j=1}^{\infty} r_j \log_2 \frac{1}{r_j} \\ &= 3.593946 \text{ bits per run.} \end{aligned}$$

Starting at the beginning of a run, suppose y is truncated after M runs of both kinds have occurred. The total number of coordinates y_n is the sum $\Sigma_1^M [R_m^{(0)} + R_m^{(1)}]$ of M independent samples each of $R^{(0)}$,$R^{(1)}$. The total random entropy is the corresponding sum $\Sigma_1^M [h(R_m^{(0)}) + h(R_m^{(1)})]$ for the samples. Omitting the detailed arguments, we obtain from the strong law of large numbers

$$H(Y) = \lim_{M\to\infty} \frac{\sum_1^M [h(R_m^{(0)}) + h(R_m^{(1)})]}{\sum_1^M [R_m^{(0)} + R_m^{(1)}]} \quad \text{w.p.1}$$

$$= \frac{H(R^{(0)}) + H(R^{(1)})}{E\{R^{(0)}\} + E\{R^{(1)}\}}$$

$$= 0.699243 \text{ bit per letter.}$$

As a check, note

$$P\{y_0 = 1\} = \frac{E\{R^{(1)}\}}{E\{R^{(0)}\} + E\{R^{(1)}\}} = \frac{5}{8},$$

which is clear from Fig. 3. The entropy of y_0 is $H(y_0) = h(3/8) = 0.954434$, and the difference

$$h(3/8) - H(Y) = H(y_0) - H(y_0|\cdots,y_{-1})$$
$$= I(y_0,\{\cdots,y_{-1}\})$$
$$= 0.255191 \text{ bit per letter}$$

is the amount by which Y fails to be a Bernoulli process. The ideal bound is

$$H(Y) \geqq 1 - h(1/8)$$
$$= 0.456436 \text{ bit per letter.}$$

The bound of Section II is easily worked out to be

$$H(Y) \geqq 1 - H(e_0|\cdots,x_{-1})$$
$$= 1 - (1/2)h(1/4)$$
$$= 0.594361 \text{ bit per letter.}$$

The entropy rate $H(E)$ of the errors can also be obtained from run-length considerations. Indeed, $\{e_n = 1\}$ is just the event $\{x_n = 0$ is a 0-run of length $1\}$ in process X. The 0-run lengths $S^{(0)}$ and the 1-run lengths $S^{(1)}$ in process X each have the geometric distribution

$$P\{S^{(0)} = j\} = P\{S^{(1)} = j\} = \frac{1}{2^j}, \quad j = 1,2,\cdots,$$

as is well known. Let the run lengths after an occurrence of $\{S^{(0)} = 1\}$ be $S_1^{(1)}, S_1^{(0)}, S_2^{(1)}, S_2^{(0)}, \cdots$, and let random variable J be the smallest $\nu \geqq 1$ for which $S_\nu^{(0)} = 1$. Since $P\{S^{(0)} = 1\} = P\{S^{(0)} > 1\} = 1/2$, we again have (1/2, 1/2) Bernoulli trials, i.e.,

$$P\{J = j\} = \frac{1}{2^j}, \quad j = 1,2,\cdots.$$

The number of intervening x_n's is the 0-run length $V^{(0)} = S_1^{(1)} + S_1^{(0)} + \cdots + S_{j-1}^{(0)} + S_j^{(1)}$ in the E process. The generating function for each $S^{(1)}$ is

$$\sum_1^\infty x^j P\{S^{(1)} = j\} = \frac{x}{2 - x},$$

and the generating function for $S^{(0)}$ conditional on $S^{(0)} > 1$ is

$$\sum_2^\infty x^j P\{S^{(0)} = j | S^{(0)} > 1\} = \frac{x^2}{2 - x},$$

so we have

$$\sum_1^\infty x^k P\{V^{(0)} = k\} = \sum_{j=1}^\infty \frac{1}{2^j} \left(\frac{x}{2 - x}\right)^j \left(\frac{x^2}{2 - x}\right)^{j-1}$$
$$= \frac{x(2 - x)}{8 - 8x + 2x^2 - x^3}, \quad |x| \leqq 1 < 1.13968.$$

Taking $(d/dx)_{x=1}$ gives $E\{V^{(0)}\} = 7$, and since the 1-runs in E have length $V^{(1)} = 1$ w.p.1, we have the check

$$P\{e_n = 1\} = \frac{E\{V^{(1)}\}}{E\{V^{(0)}\} + E\{V^{(1)}\}} = \frac{1}{8}.$$

For numerical evaluation of $H(V^{(0)})$, we use the recurrence

$$v_k = v_{k-1} - \frac{1}{4} v_{k-2} + \frac{1}{8} v_{k-3}, \quad k \geqq 4;$$

$$v_1 = \frac{1}{4}, \quad v_2 = \frac{1}{8}, \quad v_3 = \frac{1}{16}$$

satisfied by $v_k = P\{V^{(0)} = k\}$, $k \geqq 1$. The numerical result is

$$H(V^{(0)}) = \sum_1^\infty v_k \log_2 \frac{1}{v_k}$$
$$= 4.061168 \text{ bits per run},$$

giving

$$H(E) = \frac{H(V^{(0)}) + H(V^{(1)})}{E\{V^{(0)}\} + E\{V^{(1)}\}} = \frac{1}{8} H(V^{(0)})$$
$$= 0.507646 \text{ bit per letter}$$

as the entropy rate of the errors. The entropy of e_0 being $H(e_0) = h(1/8) = 0.543564$ bit per letter, the difference

$$
\begin{aligned}
h(\epsilon) - H(E) &= H(e_0) - H(e_0|\cdots,e_{-1}) \\
&= I(e_0,\{\cdots,e_{-1}\}) \\
&= 0.035918 \text{ bit per letter}
\end{aligned}
$$

is the amount by which E fails to be Bernoulli.

IV. ACKNOWLEDGMENTS

The author wishes to thank Aaron D. Wyner for bringing the problem to his attention. A paper by Berger and Lau[2] came to the author's attention after the present paper was written. Some of the results overlap; the methods are different.

REFERENCES

1. T. Berger, *Rate Distortion Theory*, Englewood Cliffs N.J.: Prentice-Hall, 1971.
2. T. Berger and J. K.-Y. Lau, "On Binary Sliding Block Codes," IEEE Trans. Inform. Theory, *IT-23*, No. 3 (May 1977).
3. P. Billingsley, *Ergodic Theory and Information*, New York: John Wiley, 1965.
4. R. M. Gray, D. L. Neuhoff, and D. S. Ornstein, *Nonblock Source Coding With a Fidelity Criterion*, Ann. Probability, *3* (1975) pp. 478–491.